Clustering human perception of environment impact using Rough Set Theory

Ani Apriani
dept. Geology Enginering
Sekolah Tinggi Teknologi Nasional
Yogyakarta, Indonesia
aniapriani@sttnas.ac.id

Iwan Tri Riyadi Yanto
dept. Information System
Ahmad Dahlan University
Yogyakarrta, Indonesia, Country
yanto.itr@is.uad.ac.id

Septiana Fathurrohmah
dept. Urban and Regional Planning
Sekolah Tinggi Teknologi Nasional
Yogyakarta, Indonesia
septiana.sttnas.ac.id

Sri Haryatmi
dept. Mathematic
Universitas Gajah Mada
Yogyakarta, Indonesia
s.kartiko@yahoo.com

Danardono
dept. Mathematic
Universitas Gajah Mada
Yogyakarta, Indonesia line
danardono@ugm.ac.id

***Abstract*— Yogyakarta, as the main destination for education, cultural exchange, and tourism in Indonesia has experienced a high demand on physical development. Despite its positive effect, the growth of physical development in Yogyakarta also brings several negative effects. This research aims to figure out the various effect of physical development in Yogyakarta based on the perception of the local residence. To achieve the objective, this research uses two methods based on rough set theory, that are Maximum dependency attribute (MDA) and fuzzy partition based on indiscernible relation (FKP). The results show that the water quality is the important attribute on physical and chemical aspects. Furthermore, on economic aspect, the highest attributes are immigration and employee absorption.**

Keywords—clustering, environment impact, Rough Set Theory

I. INTRODUCTION

Development aims to enhance the livelihood of the society by utilizing the available natural and environmental resources. However, the development itself may cause both positive and negative effects to the society. The magnitude of the development effect is influenced by the type of the policy to be implemented, characteristics of the natural environment, the availability of technology and financial resource, and the level of participation in the development process[1]–[5].

Furthermore, the effect of the spatial aspect on the development could not be neglected. This aspect may determine the effect of the development. The effect of development in the urban area, sub-urban area, and rural-area may differ due to the distinct characteristics of the type of economic activity, the intensity of urban development, and the characteristic of the social and natural environment[6], [7].

Yogyakarta, as the main destination for education, cultural exchange, and tourism in Indonesia has experienced a high demand on physical development. It has a potential market for the investment in the trading and hospitality sectors. The hospitality sector itself may contribute 24,88% in the Gross Domestic Product (GDP) of Yogyakarta. This indicate the positive impact of hospitality sector for the economic development in Yogyakarta.

Despite its positive effect, the growth of hospitality sectors in Yogyakarta also brings several negative effects. The growth of hospitality industry in Yogyakarta has caused a high demand on hotel development. The massive hotels development in Yogyakarta is said to cause several negative effects, for instance, the increase of traffic congestion, the increase of urban temperature, the decline of the groundwater reservoir, and the marginalization of the local architecture [8]. Therefore, this research aims to figure out the various effect of physical development in Yogyakarta based on the perception of the local residence. Thus, a strategic program can be well planned during their period of studies in an institution[3], [9], [10]. An effective way to detect the most its impact is the use of data mining technique[11]. Data mining in general is the process of finding, analyzing a new information that may exist in data and summarizing the results as useful information. There are many outstanding studies on data mining in the areas of clustering, association rules, classification, conflict analysis, and etc [12]–[15]. The field of data mining that concerned with the applications.

In order to achieve the research objective, two method based on rough set theory are used, that are Maximum dependency attribute (MDA) and fuzzy partition based on indiscernible relation (FKP). The MDA method is used to ranking the dependency degrees of attribute and select the most important attribute or features[16]. Furthermore, the FKP as a clustering technique is used for grouping the data selected to the class having similar characteristic[17]. Selecting, identifying the most influential attribute and grouping the data could the police maker to design the proper intervention and take immediate action to improve the quality of social environment.

II. METHOD

A. Maximum Dependency Attribute

Maximum dependency attribute (MDA) is proposed by herawan et al, the technique for selecting attribute based on the rough set theory. It uses the dependency of attribute by select the highest dependency respect to all attribute. There are three steps in MDA technique i.e : compute the equivalence classes of attribute, Determine the dependency

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE
978-1-5386-8402-3/18/$31.00 ©2018 IEEE

attribute, select the attribute based on the degree of dependency [18]–[20].

Equivalence relation can induce a unique partition. The beginning point for formatting partitions is the indistinguishable relationship denoted $IND(B)$. Let $S(U, A, V, f)$ is an information system. Two object $x, y \in U$ and attribut $B \subseteq A$ is $IND(B) \Leftrightarrow f(x,a) = f(y,a), \forall a \in A$. The equivalence classes U/B is the partition of U by $IND(B)$ and the U/B contains $x \in U$ is denoted $[x]_B$. Let $X \subseteq U$, $\underline{B}(X) = \{x \in U | [x]_B \subseteq X\}$ is the lower approximation called positive region in rough set. Then, the degree dependency D on C where $C, D \subseteq A, C \Rightarrow_k D$ is defined as

$$k = \frac{Pos_C}{|U|}, 0 \leq k \leq 1,$$

$$Pos_C = \sum_{X \in U/D} |\underline{C}(X)|.$$

B. Fuzzy partition based on indiscernible relation

Fuzzy partition based on indiscernible relation has been proposed by yanto et all. The technique is a parametric approach using likelihood of multivariate distribution function based on indiscernible relation of rough set[21]. The objective function of the technique is maximize the function $J_m(\mu, \lambda)$.

$$J_m(\mu, \lambda) = \sum_{j=1}^{J} \sum_{k=1}^{K} \mu_{jk}^m \sum_{l=1}^{I} \ln(\lambda_{\underline{a_j}}^{kl})$$

by the constrains:

$$\mu_{jk} \geq 0, \sum_{k=1}^{K} \mu_{jk} = 1,$$

$$\lambda_{\underline{a_j}}^{kl} \geq 0, \sum_{l=1}^{I} \lambda_{\underline{a_j}}^{kl} = 1,$$

By solving the nonlinear equation system of the first derivate of the Lagrange J_m, the solution of the problem are found, i.e :

$$\lambda_{U/a_j}^{kl} = \frac{\sum_{i=1}^{n} [x_i \in a_j = y_j] \mu_{ik}^m}{\sum_{i=1}^{n} \mu_{ik}^m}, \text{ for } l = 1, 2, \ldots, I_j$$

$$\mu_{jk} = \left[\sum_{s=1}^{K} \left[\frac{\sum_{l=1}^{I} \ln(\lambda_{U/a_j}^{kl})}{\sum_{l=1}^{I} \ln(\lambda_{U/a_j}^{sl})} \right]^{\frac{1}{m-1}} \right]^{-1}$$

III. RESEARCH METHOD

This research aims to grouping the city having similar impact of environment dataset. There are two phases of the process. Firstly, the selecting the most important of the attribute of the dataset by ranked the degree dependency based on the MDA technique. Size of the dataset are reduced by excluded attribute having the less degree of dependency. Then, the reduced size dataset is grouped using the clustering technique named Fuzzy partition based on indiscernible relation. This can help the police maker to design the proper intervention and take immediate action to improve the quality of social environment.

The primer dataset was collected by survey involved the 400 respondents consists of 224 females and 176 males. The respondents are fill the questioner regarding the perception of the environment impact in term of Physic and chemical aspects and Economic aspect. The reliability score of the data is 0.953 in Cronbach alpha test. Table 1. Shows the description of the data set.

TABLE I. DATA DESCRIPTION

Data set	Attribute	Description
Physic and chemical aspects	water quantity, water quality, water absorption level, temperature, air pollution level, climate, noise level, land use, availability of public open space	Number of attribute: 9 Mean : 25 SD : 2.95
Economic aspect	immigration, rate of employment, economic structure development, revenue, expenditure, shift of occupation, public health, increasing number of educational facility, increasing number of religious facility, increasing number of health care facility	Number of attribute: 10 Mean : 16 SD : 2.67

IV. RESULTS AND DISCUSSION

A. Physic and chemical aspects

Physic and chemical aspects consists nine attributes with mean = 25 and SD = 2.95. The Degree of dependency is shown in table 2. From Table 2, attribute L3, L4, L6, L7, L8 and L9 are the less of dependency thus the attributes are excluded so that the L1, L2 and L5 are selected as an attributes to the next phases. Moreover, the high dependency is L2 that is the water quality as an important attribute.

B. *Economic aspect*

Economic aspect consists ten attributes with mean = 12 and SD=1.59. The Degree of dependency id shown in table 3. From Table 3, The highest attributes are E1 (immigration), the movement of the population from outside into the city of Yogyakarta and E2 (Employee absorption). Attribute E3, E7, E8 and E10 are the less of dependency thus the attributes are excluded so that the E1, E2, E4, E5 and E6 are selected as an attributes to the next phase.

TABLE II. DEGREE DEPENDENCY OF PHYSIC AND CHEMICAL ASPECTS

Attribute	L1	L2	L3	L4	L5	L6	L7	L8	L9
Degree of dependency	0.0075	0.0475	0.0025	0.0025	0.0075	0.0025	0.0025	0.0025	0.0025

TABLE III. DEGREE DEPENDENCY OF ECONOMIC ASPECT

Attribute	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
1stDegree of dependency	1	1	0	1	1	0.015	0	0	0	0
2ndDegree of dependency	0.01	0.01	0	0	0	0	0	0	0	0

C. *Cluster Analysis*

The cluster result is analyzed using silhouette coefficient to determine the desirable number of cluster. In the clustering phase, the degree of fuzziness is selected as $1.2 \leq m \leq 2$ with the variation number of clusters are 2 until 14. The average of silhouette coefficients is summarized in table 4. From table 4. Can be shown that the number of clusters can be determine between 10 -14 number of cluster with significantly value of silhouette coefficients rising 0.9.

TABLE IV. SILHOUETTE COEFFICIENTS IN VARIOUS NUMBER OF CLUSTERS

Number of cluster	Physic and chemical	Economic aspects
2	0.487	0.522
3	0.551	0.614
4	0.632	0.681
5	0.691	0.778
6	0.761	0.810
7	0.830	0.857
8	0.871	0.890
9	0.902	0.909
10	0.925	0.940
11	0.937	0.956
12	0.938	0.973
13	0.954	0.982
14	0.961	0.995

V. CONCLUSION

This paper studies the utilization of the techniques to cluster the environment impact dataset. The technique consists two phase i.e feature selection based on the maximum dependency attribute and then clustering using fuzzy partition based on indiscernible relation. Two aspect of environment impact dataset are used to be utilizing using the both techniques. the usefulness of the technique has been demonstrated as an experiment results. This paper can be used to make recommendation to improve the quality of social environment.

ACKNOWLEDGMENT

This work was supported by Directorate of Research and Community Service Kemenristekdikti no SK 58/STTNAS/P3M/Pen.Dikti/V 12017

REFERENCES

[1] P. Clavel and R. Young, "'Civics': Patrick Geddes's theory of city development," *Landsc. Urban Plan.*, vol. 166, no. June, pp. 37–42, 2017.

[2] A. Kumari and A. K. Sharma, "Physical & social infrastructure in India & its relationship with economic development," *World Dev. Perspect.*, vol. 5, pp. 30–33, 2017.

[3] P. Ashcroft and L. Murphy Smith, "Impact of environmental regulation on financial reporting of pollution activity: A comparative study of U.S. and Canadian firms," *Res. Account. Regul.*, vol. 20, pp. 127–153, 2008.

[4] J. K. Woo, D. S. H. Moon, and J. S. L. Lam, "The impact of environmental policy on ports and the associated economic opportunities," *Transp. Res. Part A Policy Pract.*, no. xxxx, pp. 0–1, 2017.

[5] P. Ashcroft and L. Murphy Smith, "Impact of environmental regulation on financial reporting of pollution activity: A comparative study of U.S. and Canadian firms," *Res. Account. Regul.*, vol. 20, pp. 127–153, Jan. 2008.

[6] K. Howard and R. Gerber, "Impacts of urban areas and urban growth on groundwater in the Great Lakes Basin of North America," *J. Great Lakes Res.*, vol. 44, no. 1, pp. 1–13, 2018.

[7] A. R. Shahtahmassebi *et al.*, "How do modern transportation projects impact on development of impervious surfaces via new urban area and urban intensification? Evidence from Hangzhou Bay Bridge, China," *Land use policy*, vol. 77, pp. 479–497, 2018.

[8] P. K. Yogyakarta, "DAMPAK PERTUMBUHAN HOTEL TERHADAP PERUBAHAN KARAKTERISTIK PERWILAYAHAN KOTA YOGYAKARTA Tika Ainunnisa Fitria," vol. 10, pp. 52–57, 2016.

[9] M. T. Dugan, E. H. Turner, M. A. Thompson, and S. M. Murray, "Measuring the financial impact of environmental regulations on the trucking industry," *Res. Account. Regul.*, vol. 29, no. 2, pp. 152–158, 2017.

[10] C. Mary Schooling, E. W. L. Lau, K. Y. K. Tin, and G. M. Leung, "Social disparities and cause-specific mortality during economic development," *Soc. Sci. Med.*, vol. 70, no. 10, pp. 1550–1557, 2010.

[11] T. Woldai and A. G. Fabbri, "The Impact of Mining on The Environment," in *Deposit and Geoenvironmental Models for Resource Exploitation and Environmental Security*, A. G. Fabbri, G. Gaál, and R. B. McCammon, Eds. Dordrecht: Springer Netherlands, 2002, pp. 345–364.

[12] R. Wardoyo, "A Review on fuzzy multi-criteria decision making land clearing for oil palm plantation," vol. 1, no. 2, pp. 75–83, 2015.

[13] I. Alwiah, "RETRACTED•: RCE-Kmeans Method for Data Clustering," p. 26555, 2017.

[14] M. Muhajir and B. Rian, "Association Rule Algorithm Sequential Pattern Discovery using Equivalent Classes (SPADE) to Analyze the Genesis Pattern of Landslides in Indonesia," vol. 1, no. 3, pp. 158–164, 2015.

[15] D. Ismi, S. Panchoo, and M. Murinto, "K-means clustering based filter feature selection on high dimensional data," *Int. J. Adv. Intell. Informatics*, vol. 2, no. 1, pp. 38–45, 2016.

[16] N. Senan, R. Ibrahim, N. M. Nawi, I. T. R. Yanto, and T. Herawan, *Soft set theory for feature selection of traditional Malay musical instrument sounds*, vol. 6377 LNCS, no. M4D. 2010.

[17] I. T. R. Yanto, M. A. Ismail, and T. Herawan, "A modified Fuzzy k-Partition based on indiscernibility relation for categorical data clustering," *Eng. Appl. Artif. Intell.*, vol. 53, pp. 41–52, 2016.

[18] T. Herawan, I. T. R. Yanto, and M. Mat Deris, *Rough set approach for categorical data clustering*, vol. 64. 2009.

[19] T. Herawan, M. M. Deris, and J. H. Abawajy, "A rough set approach for selecting clustering attribute," *Knowledge-Based Syst.*, vol. 23, no. 3, pp. 220–231, Apr. 2010.

[20] N. Senan, R. Ibrahim, N. Mohd Nawi, I. T. R. Yanto, and T. Herawan, *Rough set approach for attributes selection of traditional Malay musical instruments sounds classification*, vol. 151 CCIS, no. PART 2. 2011.

[21] I. T. R. Yanto, M. A. Ismail, and T. Herawan, "A modified Fuzzy k-Partition based on indiscernibility relation for categorical data clustering," *Eng. Appl. Artif. Intell.*, vol. 53, pp. 41–52, Aug. 2016.